Los Alamos NATIONAL LABORATORY EST.1943

LA-UR-15-20576 (Accepted Manuscript)

High-Order / Low-Order Methods for Ocean Modeling

Newman, Christopher Kyle
Womeldorff, Geoffrey Alan
Chacon, Luis
Knoll, Dana Alan

Provided by the author(s) and the Los Alamos National Laboratory (2015-12-21).

To be published in: Procedia Computer Science, Vol. 51, pp. 2086-2096, 2015

DOI to publisher's version: 10.1016/j.procs.2015.05.477

Permalink to record: http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-15-20576

Disclaimer:
Approved for public release. Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

RESEARCH LIBRARY

Procedia Computer Science

Volume 51, 2015, Pages 2086–2096

ICCS 2015 International Conference On Computational Science

High-Order / Low-Order Methods for Ocean Modeling

Christopher Newman[1], Geoff Womeldorff[2], Luis Chacón[3], and Dana A. Knoll[4]

[1] Fluid Dynamics and Solid Mechanics Group (T-3), Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A. cnewman@lanl.gov

[2] Applied Computer Sciences Group (CCS-7), Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A. womeld@lanl.gov

[3] Applied Mathematics and Plasma Physics Group (T-5), Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A. chacon@lanl.gov

[4] Integrated Design and Assessment Group (XTD-IDA), Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A. nol@lanl.gov

Abstract

We examine a High Order / Low Order (HOLO) approach for a z-level ocean model and show that the traditional semi-implicit and split-explicit methods, as well as a recent preconditioning strategy, can easily be cast in the framework of HOLO methods. The HOLO formulation admits an implicit-explicit method that is algorithmically scalable and second-order accurate, allowing timesteps much larger than the barotropic time scale. We show how HOLO approaches, in particular the implicit-explicit method, can provide a solid route for ocean simulation to heterogeneous computing and exascale environments.

Keywords: HOLO methods, ocean modeling, IMEX method, nonlinear elimination

1 Introduction

The High-Order / Low-Order (HOLO) approach is a moment-based, scale-bridging algorithm where the coarse scale (LO) problem is obtained via moment integration and is used to accelerate the fine scale (HO) problem. The dimensionality of this LO problem is significantly smaller than the HO problem and, therefore, the LO system is far less expensive to solve. The HOLO approach provides multigrid-like algorithmic acceleration: the LO problem solver relaxes long wavelength components of the solution, while the HO problem solver relaxes short wavelength components. These methods accelerate solution convergence by alternating the solution of the HO system and a LO system, which are forced to remain discretely consistent (down to the level of truncation error). The algorithmic idea builds on a well-defined hierarchical description (moments) of widely varying space and time scales.

The hierarchical nature of the algorithm lends itself readily to emerging architectures. The algorithm exploits an isolation of scales, thus embracing multilevel parallelism and asynchrony. Such scale isolation can allow for maximizing flops with minimal data movement, thus providing

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2015
© The Authors. Published by Elsevier B.V.

doi:10.1016/j.procs.2015.05.477

a key advantage as communication costs continue to dominate relative to floating point costs in emerging architectures.

The HOLO approach has been very successful in plasma simulation [5, 4, 27], neutral partical transport [16, 22], and thermal radiative transfer problems [23, 24]. In addition, the HOLO concept has played a critical role in the co-design of numerical algorithms and their supporting exascale implementations [1].

Accurate modeling of ocean circulation under various forcings is of great importance, and impacts a variety of political decisions. In particular, it has been shown that high resolution is necessary for an accurate simulation of long-time ocean thermohaline circulation [18]. Scientists are currently performing ocean simulations with semi-implicit [6] and split-explicit methods [9, 11]. Typically, these methods utilize a barotropic-baroclinic splitting that isolates fast (barotropic) and slow (baroclinic) time scales. The faster external gravity waves (or barotropic motions) are independent of depth, and thus two dimensional, while the slower baroclinic motions are fully three dimensional. For most problems of interest, explicit time discretizations are impractical for these systems, due to short timesteps imposed by the fast waves.

Methods based on barotropic-baroclinic splitting can easily be cast in the framework of HOLO methods. Indeed, the semi-implicit (SI) method utilized in the Parallel Ocean Program (POP) [6, 26] and split-explicit (SE) methods utilized in Model for Prediction Across Scales (MPAS) [25] and Modular Ocean Model (MOM) [21] are characterized by a LO barotropic system obtained by vertical moment of the HO baroclinic continuity and momentum equations.

Recently, HOLO has been exploited as a preconditioner for fully-coupled, fully-implicit ocean models. Fully implicit approaches have been sought to remove timestep stability restrictions, enhance numerical accuracy, and by the need for high spatial resolution. One advantage of these methods is that relatively large timesteps can be taken without sacrificing second-order accuracy. A disadvantage, however, is that fully-implicit methods require a non-linear solution for each timestep. The strategy outlined in [19] effectively preconditions the fully coupled HO system by only inverting the LO system. Algorithmic scalability is afforded by multigrid methods.

Another approach is that of an implicit-explicit (IMEX) methods [3, 7, 14]. This approach is motivated by the successful implementation of an iterated IMEX method for radiation hydrodynamics problems [12, 13], related work on sea-ice modeling [17] and plasma physics [5, 4]. In this particular class of IMEX methods, the two dimensional scalar continuity LO equation is treated implicitly with preconditioned JFNK, and the remaining three dimensional HO equations are driven by the LO barotropic solution and subcycled explicitly within the JFNK residual evaluation. Thus, the fast barotropic physics are treated implicitly while the slower baroclinic physics are treated explicitly. When implemented appropriately, the HOLO approach allows for an algorithmically scalable, second-order time integration, with timesteps much larger than methods with timesteps restricted by the barotropic time scale.

Explicit treatment and subcycling of the HO problem allow larger LO timesteps to be taken, and also allow a avenue of communication minimization (see §3.4). Therefore, HOLO approaches, in particular the IMEX method, can provide a solid route for ocean simulation to heterogeneous computing and exascale environments by minimizing communication [20], since communication costs continue to grow and are becoming dominant relative to floating point costs (see §3.4).

The manuscript is organized as follows: we present a mathematical formulation of HOLO for a z-level ocean model in §2, draw attention to some specific HOLO approaches for ocean modeling in §3, and make comparison of some of the recent approaches in terms of accuracy,

efficiency and parallel implementation in §4. Conclusions are drawn in §5.

2 HOLO formulation for a z-level ocean model

We assume a z-level ocean model in Cartesian coordinates, in K layers, where $k = 1$ corresponds to the top layer [6]. The HO equations (momentum and continuity) are given by

$$\frac{\partial \mathbf{u}_k}{\partial t} + L_1\left(\mathbf{u}_k, \eta\right) + f\mathbf{u}_k^{\perp} + G_1\left(\varphi_{i,k}, \eta\right) + G_2\left(\varphi_{i,k}\right) = 0, \qquad k = 1, \ldots, K, \tag{1}$$

$$\frac{\partial \eta}{\partial t} + D(\mathbf{u}_1, \eta) = 0, \qquad k = 1, \quad \text{and} \tag{2}$$

$$\nabla \cdot (h_k(\eta)\mathbf{u}_k) + w_{k-1/2} - w_{k+1/2} = 0, \qquad k = 2, \ldots, K, \tag{3}$$

with

$$L_1\left(\mathbf{u}_k, \eta\right) = \left(\mathbf{u}_k \cdot \nabla\right)\mathbf{u}_k + \frac{w_{k-1/2}\mathbf{u}_{k-1/2} - w_{k+1/2}\mathbf{u}_{k+1/2}}{h_k(\eta)}, \tag{4}$$

$$G_1\left(\varphi_{i,k}, \eta\right) = \frac{1}{\rho_0} g \nabla \rho_1(\varphi_{i,k})\eta, \tag{5}$$

$$G_2\left(\varphi_{i,k}, \eta\right) = \frac{1}{\rho_0}\nabla p_{H\,k}(\varphi_{i,k}) \quad \text{and} \tag{6}$$

$$D(\mathbf{u}_1, \eta) = \nabla \cdot (h_1(\eta)\mathbf{u}_1) - w_{3/2}, \tag{7}$$

where the vertical discretization consists of equally spaced layers of thickness Δz. Here $\mathbf{u}_k = [u_k \; v_k]^{\mathrm{T}}$ and p_k are the horizontal velocity and pressure in layer k, w_k is the vertical velocity at the bottom of layer k, $w_{k-1/2}$ is the vertical velocity at mid-layer k, η is sea surface height perturbation, h_k is the thickness of layer k with $h_k = \Delta z$ for $k \neq 1$ and $h_1 = \Delta z + \eta$, f is the Coriolis parameter, $\rho_0 = 1000$ is the reference density, t corresponds to time, and ∇ and $\nabla\cdot$ are the gradient and divergence operators in the horizontal plane. We augment (1)–(3) with transport equations for temperature $\varphi_{1,k}$, and salinity $\varphi_{2,k}$ in each layer k:

$$\frac{\partial}{\partial t}(h_k(\eta)\varphi_{i,k}) + L_2(\varphi_{i,k}, \mathbf{u}_k, \eta) = 0, \qquad i = 1, 2, \tag{8}$$

with

$$L_2(\varphi_{i,k}, \mathbf{u}_k, \eta) = \nabla \cdot (h_k(\eta)\varphi_{i,k}\mathbf{u}_k) + \frac{w_{k-1/2}h_{k-1/2}(\eta)\varphi_{i,k-1/2} - w_{k+1/2}h_{k+1/2}(\eta)\varphi_{i,k+1/2}}{h_k(\eta)}, \tag{9}$$

and a linear equation of state for density given by

$$\rho_k(\varphi_{1,k}, \varphi_{2,k}) = \rho_0(1.0 - \alpha\,\varphi_{1,k} + \beta\,\varphi_{2,k}), \tag{10}$$

with $\alpha = 2.5 \times 10^{-4}$ and $\beta = 7.6 \times 10^{-4}$ [26]. The pressure in layer k is given by $p_k = g\,\rho_1\,\eta + \sum_{l=1}^{k} g\,\rho_l\,h_l$, where g is acceleration due to gravity. Due to (10), $p_k = p_k(\varphi_{i,k})$. We assume the fluid is initially at rest, with $\mathbf{u}_k = 0$, $k = 1, \ldots, K$ on the boundary and $w_0 = w_K = 0$. Variable definitions and associated units are listed in Table 1.

$\mathbf{u}_k$	horizontal velocity	m/s	w_k	vertical velocity	m/s
p_k	pressure	$\mathrm{N/m^2}$	h_k	layer thickness	m
η	surface perturbation	m	f	Coriolis parameter	$\mathrm{s^{-1}}$
ρ_k	density	$\mathrm{kg/m^3}$	t	time	s
$\varphi_{1,k}$	temperature	°C	$\varphi_{2,k}$	salinity	PSU

Table 1: Variable definitions and units for (1)–(3) and (8).

The fastest time scale associated with (1)–(3) and (8) is the barotropic timescale, due to the external gavity-wave, $\Delta t_g = \Delta x \, (gD)^{-1/2}$, with speed on order of 200 m/s. The baroclinic timescale, Δt_a, is due to advection or dynamic time scales with speed on order of 10 m/s, and is therefore much slower. The separation between these time scales presents difficulties for time integration. Explicit methods are generally unstable for timesteps larger than the external gravity-wave time scale.

To obtain the LO system, we introduce the vertical moment

$$\overline{\mathbf{u}} = \frac{1}{H(\eta)} \sum_{k=1}^{K} h_k(\eta) \mathbf{u}_k, \tag{11}$$

where $H(\eta) = \sum_{k=1}^{K} h_k(\eta)$ and $\overline{\mathbf{u}}$ is the LO (2-D) velocity. Application of (11) to (1)–(3) yields the LO problem

$$\frac{\partial \eta}{\partial t} + \nabla \cdot H(\eta) \overline{\mathbf{u}} = 0, \tag{12}$$

$$\frac{\partial \overline{\mathbf{u}}}{\partial t} + g \nabla \eta = \overline{G}, \tag{13}$$

where

$$\overline{G} = \frac{1}{H(\eta)} \sum_{k=1}^{K} -L_1(\mathbf{u}_k) - f \mathbf{u}_k^{\perp} - \nabla p_{H\,k}. \tag{14}$$

The LO problem constitutes a 2-D scalar (rather than a 3-D vector) system. The HO problem is given by (1)–(3) and (8). In addition, the physics associated with the external gravity-wave are isolated to the LO system and removed from the HO system. The approximate decoupling of the LO and HO systems allows for the independent time integration of each system separately, with timestep size Δt_{LO} and Δt_{HO}, respectively. We note that the LO system has a CFL condition associated with Δt_g, while the HO system has a CFL condition associated with $\Delta t_a > \Delta t_g$. We also note that, upon temporal discretization, the LO system (12)–(13) can often be reduced to an elliptic equation for η^{n+1} at the current timestep:

$$\left(\frac{1}{\tau} - \nabla \cdot H \nabla \right) \eta^{n+1} = F(\eta, \overline{\mathbf{u}}), \tag{15}$$

where τ is a function of timestep size, η, and $\overline{\mathbf{u}}$ are are at previous timesteps (see [6] for details). We note that with N horizontal mesh points, the HO problem (1)–(3) and (8) consists of approximately $5\,NK$ unknowns; while the LO problems (12)–(13) and (15) consist of $2\,N$ and N unknowns respectively. A comparison of cost for solving the HO and LO problems is given in §4.2.

3 HOLO approaches for Ocean Modeling

3.1 Semi-implicit methods

The earliest example of an HOLO approach to ocean modeling is the semi-implicit (SI) method. Typical SI methods treat the LO system implicitly using (15) with timesteps on order of the baroclinic time scale, $\Delta t_{\text{LO}} \approx \Delta t_a$, and the HO system explicitly with $\Delta t_{\text{HO}} = \Delta t_{\text{LO}}$. SI is the approach of POP [6, 26]. The implementation in POP utilizes the conjugate gradient method with Jacobi preconditioning for solution of (15) and a leapfrog scheme for time integration of the HO system. POP has been used successfully for global, high resolution (0.1°) ocean modeling [29]. The SI method has an advantage over fully-explicit methods in that timesteps larger than Δt_g can be readily utilized. In practice, SI can suffer from first-order temporal accuracy. The split-explicit method was developed to address these issues [25].

3.2 Split-explicit methods

The split-explicit (SE) algorithm is the current method used in the ocean community, e.g. in MPAS [25] and MOM [21]. In contrast to SI, SE methods treat the LO system (12)–(13) with explicit subcycling. For a single timestep of SE, the HO problem is integrated explicitly with $\Delta t_{\text{HO}} \approx \Delta t_a$ and the LO problem is explicitly subcycled with $\Delta t_{\text{LO}} \approx \Delta t_g$. In practice, the HO and LO systems are coupled via a Picard iteration in order to achieve second-order temporal accuracy [9, 11]. For each iteration, the HO coupling term is provided as a forcing term for the LO problem. The LO problem is explicitly subcycled using a predictor-corrector scheme, and upon completion a LO iterate is obtained. This iterate is used to integrate the HO problem one step using the same predictor-corrector scheme to obtain a HO iterate. Upon convergence of the Picard iteration, the solution to both the LO and HO problems at the new timestep is found.

3.3 Fully-implicit methods with HOLO as a Preconditioner

The HOLO formulation has recently been utilized as a preconditioner for a fully implicit, second-order time integration of (1)–(3) and (8). In this approach, the nonlinear system is solved using a Jacobian-free Newton-Krylov (JFNK) method [15], with a physics-based preconditioner that consists only of the LO system [19]. The JFNK framework allows tighter coupling of the physics, thus reducing numerical errors and improving stability versus operator-splitting techniques.

The key to an efficient implementation of JFNK is effective preconditioning. In particular, in [19] a physics-based preconditioner strategy was developed for ocean simulation based on the HOLO formulation. The strategy outlined in [19] effectively preconditions the HO system by only inverting the LO system (15). In [19], it is shown that (15) can be approximately inverted in a scalable way using multigrid methods [28]. Specifically, the implementation in [19] utilizes the Trilinos NOX package for application of the JFNK method and the Trilinos ML [10] package for approximate inversion of (15). The preconditioner is demonstrated to allow a stable time integration with timesteps $\Delta t_{\text{HO}} \approx \Delta t_a$ while greatly reducing the number of GMRES iterations required per Newton iteration.

3.4 Implicit-explicit methods

In the implicit-explicit (IMEX) approach, the LO problem (12) is treated implicitly with preconditioned JFNK and $\Delta t_{\text{LO}} > \Delta t_a$, and the HO problem is driven by the LO solution and

subcycled explicitly within the JFNK residual evaluation with $\Delta t_{\text{HO}} \approx \Delta t_a$. This approach is known as *nonlinear elimination*, since the HO equations are eliminated from the the implicit nonlinear residual. After this nonlinear elimination, the resulting nonlinear system consists only of the LO equation [5, 30].

For a single timestep of IMEX, a nonlinear LO problem (12) must be solved. The JFNK solution procedure requires that the LO problem be evaluated for each GMRES iteration. To evaluate the LO problem, the HO problem is explicitly subcycled with a predictor-corrector scheme given the current LO iterate such that a suitable second-order velocity is obtained. The LO velocity $\overline{\mathbf{u}}$ is computed from the HO solution via (11) and provided to the LO problem. Upon convergence of JFNK, a solution to both the LO and HO problems at the new timestep is obtained. We note that the subcycling process is executed for every linear iteration, and thus effective preconditioning is a vital requirement to minimize the total number of function evaluations per LO timestep. The preconditioner for the LO system is formulated as (15), hence the same preconditioner as in [19] is used (see §3.3). The associated Krylov vector and nonlinear machinery reside in the LO space. Hence, IMEX requires a smaller memory footprint over fully-implicit methods.

In the SE method, the HO system is advanced with a long timestep, Δt_{HO}, and the LO system is subcycled with short timesteps, Δt_{LO}, restricted by a barotropic CFL limit, within a single HO timestep. Hence, $\Delta t_{\text{LO}} < \Delta t_{\text{HO}}$ necessarily. The SI method utilizes one implicit LO timestep within the HO system, thus $\Delta t_{\text{LO}} = \Delta t_{\text{HO}}$. The IMEX approach differs from the traditional SE and SI approaches because the LO system is advanced implicitly and the HO system is explicitly subcycled within the LO system. This allows $\Delta t_{\text{LO}} > \Delta t_{\text{HO}}$ while providing a stable time integration strategy.

In parallel, $\Delta t_{\text{LO}} > \Delta t_{\text{HO}}$ provides a route to a communication-avoiding algorithm. In modern and evolving computer architectures, communication will be the dominant bottleneck. A communication staging strategy has been designed to minimize communication within the explicit HO solver by reducing the frequency of halo exchanges per timestep. The strategy increases the communication halo width from its conventional value of 1 to a value based heuristically on the amount of subcycling required. For explicit methods, as in the case of the HO solver, the required halo communication per timestep can be determined *a priori*. This allows the algorithm to perform a single communication call for the HO system per LO timestep. This technique results in significant communication savings, at the expense of performing additional computations within the extended halo.

4 Comparison of HOLO methods

We examine a fully-implicit (trapezoidal rule) method (TR-JFNK), SE, and IMEX on a prototype problem where the domain consists of a two dimensional infinite channel discretized horizontally with a C-grid finite volume scheme [2] with $\Delta x = 1.0 \times 10^4$, $K = 60$, $L = 5.0 \times 10^6$, $D = 5.0 \times 10^3$, $f = 1.0 \times 10^{-4}$. The fluid is initially at rest, with an initial perturbation in η. In addition, the initial condition for temperature is a linear distribution in depth with $\varphi_{1,1} = 20.0$ and $\varphi_{1,60} = 10.0$, combined with a perturbation in layers near the surface. The initial condition for salinity is a linear distribution in depth with $\varphi_{2,1} = 30.0$ and $\varphi_{2,60} = 26.0$.

4.1 Fully-implicit method with HOLO preconditioner

Figure 1a shows convergence as a function of timestep size for TR-JFNK, preconditioned by direct LU inversion (TR-JFNK-LU, dashed line), explicit 4th order Runge–Kutta (RK4, dashed-

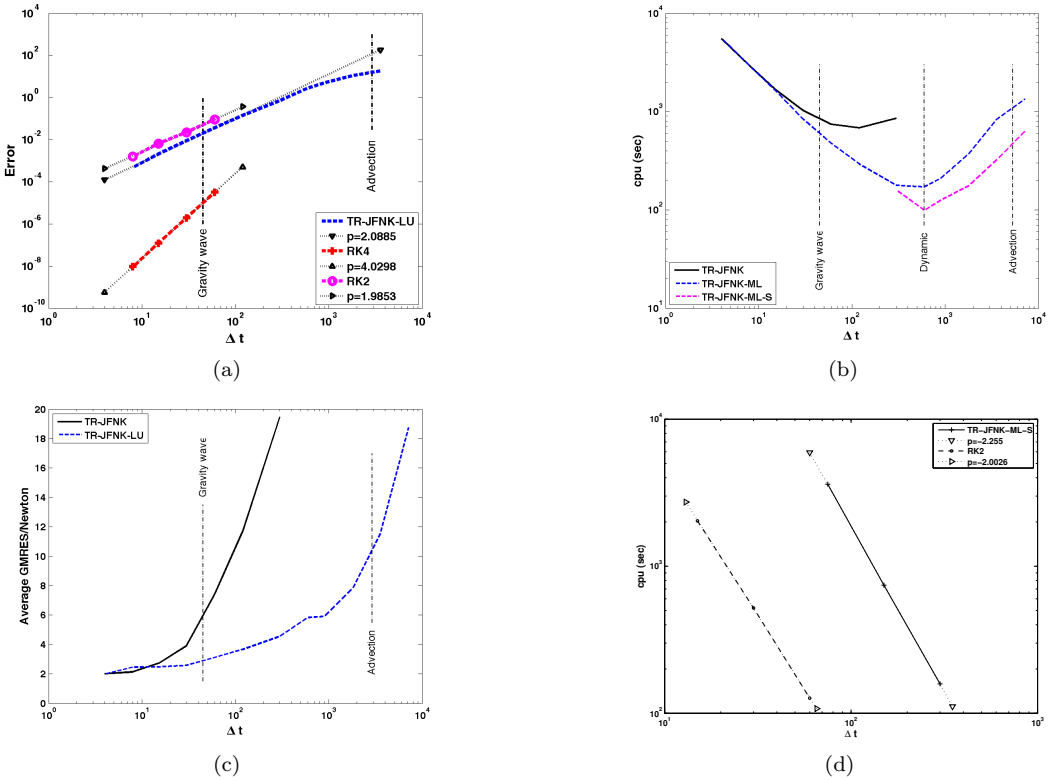

Figure 1: (a) Timestep size vs. convergence. (b) Timestep size vs. CPU. (c) Timestep size vs. average number of GMRES iterations per Newton iteration. (d) Timestep size vs. CPU; scalability with respect to problem size.

cross line), and 2nd order Runge–Kutta (RK2, dashed-dotted line). Convergence is computed relative to the RK4 solution with $\Delta t = 1.875$. Also shown is an exponential fit to the convergence that clearly shows TR-JFNK-LU and RK2 to have second-order convergence and RK4 to have fourth-order convergence.

Figure 1b shows cost in CPU seconds as a function of timestep size for unpreconditioned TR-JFNK (solid line), and TR-JFNK with application of preconditioner by the Trilinos algebraic multigrid package ML (TR-JFNK-ML, dotted line), and one SIMPLE sweep [8] (TR-JFNK-ML-S, dashed-crossed line). The figure shows that, without preconditioning, the implicit method only scales well with timesteps up to Δt_g, while effective preconditioning allows ideal algorithmic scaling with timesteps on order Δt_a.

Figure 1c shows the average number of GMRES iterations per Newton iteration as a function of timestep size for TR-JFNK (solid line) and TR-JFNK-LU (dashed line). While the average number of GMRES iterations per Newton iteration steadily increases for TR-JFNK, the number of iterations increases more rapidly for timesteps larger than Δt_g. For TR-JFNK-LU, the average number of GMRES iterations per Newton iteration remains constant for timesteps up to Δt_g and increases modestly up to Δt_a. The advantage of preconditioning in terms of the number of GMRES iterations is clearly shown. Algorithmic scalability of the method is a result of the preconditioner greatly reducing GMRES per Newton for large timesteps.

TR-JFNK-ML-S			RK2		
Δt	Δx	# unknowns	Δt	Δx	# unknowns
75	2.5×10^3	482,060	15	2.5×10^3	482,060
150	5.0×10^3	241,060	30	5.0×10^3	241,060
300	1.0×10^4	120,560	60	1.0×10^4	120,560

Table 2: Δt, Δx, and number of unknowns.

The problem was repeated twice with finer spatial and temporal resolution to demonstrate scalability of the preconditioned implicit trapezoidal rule method. The ratio $\Delta t/\Delta x$ was fixed for each problem to respect the dynamic time scale, with Δt, Δx and problem size given in Table 2. Figure 1d shows cost in CPU seconds as a function of timestep size for the three resolutions using TR-JFNK-ML-S (solid line) and second-order RK2 (dash-dotted line). The figure clearly shows that the preconditioned implicit method CPU scales proportionally with problem size and temporal resolution.

The results show that the HOLO formulation, utilized as a preconditioner for a fully-implicit method, allows timesteps larger the gravity wave time-scale. In addition, the preconditioner requires only the solution of the LO system. The preconditioner allows the implicit method cost to scale as explicit methods, but with larger implicit timesteps, and is second-order accurate.

4.2 Split-explicit and implicit / explicit methods

The problem was repeated for SE and IMEX. Recall that subcycling of the HO component of IMEX is required such that $\Delta t_{\text{HO}} = \Delta t_a$ when $\Delta t_{\text{LO}} > \Delta t_a$, and subcycling of the LO component of SE is required such that $\Delta t_{\text{LO}} = \Delta t_g$ when $\Delta t_{\text{HO}} > \Delta t_g$. Algorithmic results are presented in Figures 2a–2b. Note that in the figures the timestep axis refers to the *outer* timestep: Δt_{LO} for IMEX and Δt_{HO} for SE.

Figure 2a shows error as a function of timestep size for IMEX (solid line) and SE (dotted line) and clearly shows IMEX and SE to have second-order convergence. Note that SE shows a transition zone between second-order slopes, which occurs at timesteps where subcycling becomes active. The transition reflects the fact that the error is dominated by the smaller (subcycling) timestep. Figure 2b shows total serial CPU time (both HO and LO components) as a function of timestep size for IMEX and SE, and shows both methods scale well, although IMEX costs more than SE over the range of timesteps. Note that the LO SE component costs more and does not scale as well as the IMEX LO; and HO is the dominant cost in both SE and IMEX. The algorithmic scability of the IMEX LO problem is achieved entirely by effective preconditioning. These results show that the IMEX algorithm is algorithmically scalable and second-order accurate, with serial performance comparable to the split-explicit method.

When IMEX is implemented with an effective communication staging strategy, cost of the IMEX HO solver in parallel can be significantly reduced. Most ocean models utilize domain decomposition in the horizontal surface, with subdomains extruded vertically in depth. Thus there is horizontal halo exchange (communication) required on each layer. Traditional single-halo implementations perform a single halo exchange once per timestep or stage. Our communication staging strategy [20] aggregates the network communications of halo cells necessary for multiple explicit updates by increasing the width of the halo band relative to the number of staged timesteps. In our implementation, the halo width is increased and the necessary halo communication for multiple timesteps is performed in advance at the beginning of the timestep, followed by subcycled timesteps. This strategy particularly favors our IMEX method over SE, due to the fact that IMEX subcycles the HO components over longer timesteps, rather than

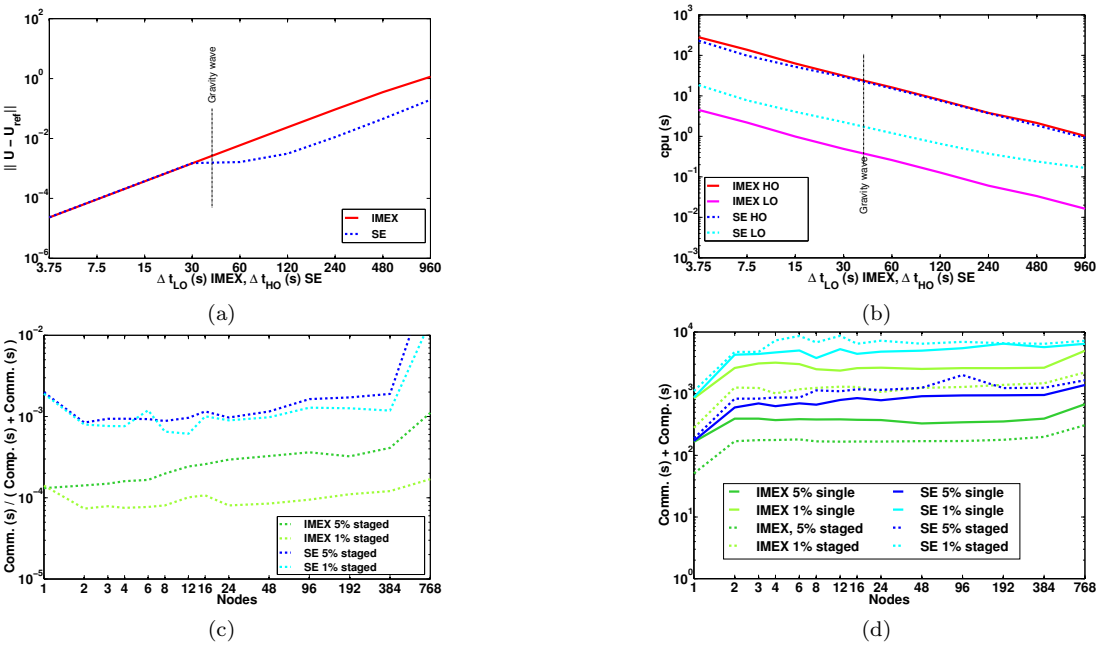

Figure 2: (a) Timestep size vs. error for IMEX and SE. (b) Timestep size vs. CPU for HO and LO components of IMEX and SE. (c) Relative percent communication. (d) Computation time.

subcycling the LO components over smaller timesteps.

Figures 2c–2d show weak scaling results to evaluate effectiveness of the communication staging strategy for both IMEX and SE, with $\Delta t_{\mathrm{HO}} = 900$ ($\Delta t_{\mathrm{LO}} = \Delta t_g$) for SE and $\Delta t_{\mathrm{LO}} = 900$ ($\Delta t_{\mathrm{HO}} = \Delta t_a$) for IMEX. With this choice of *outer* timestep size, the SE LO component must be subcycled 30 times and the IMEX HO component must be subcycled 3 times. The motivation for this problem is to compare the relative communication cost for IMEX and SE for integration at large timesteps. The computations were performed on the Mustang system at Los Alamos National Laboratory. To demonstrate network communication effects, these examples were run using one single-threaded MPI rank per node (one compute-core per node) on 1–768 nodes in two weak scaling configurations. The first configuration, denoted 5% fixes the ratio of halo width to local vector length to 1 : 9, and the second configuration, denoted 1%, fixes the ratio to 1 : 49.

Figure 2c shows that the communication staging strategy reduces the relative communication costs for IMEX by an order of magnitude over that of SE. Figure 2d shows the computation time for SE and IMEX. The figure confirms that, with computation rapidly becoming a fixed cost per method, choice of staged versus single-halo, and the halo size have significant impact on parallel performance. Figure 2d also shows that staging increased computational time for SE, while reducing that of IMEX. We note that staging ameliorates the upturn communications cost at the highest node-count, and that, in all test combinations, IMEX has a lower total time than SE.

These results show that, with an effective communication staging strategy and subcycling of the HO problem, the IMEX method in a parallel environment can decrease the relative communication by more than an magnitude over the split-explicit approach. This makes IMEX ideally suited for heterogeneous massively-parallel environments, where communication costs

continue to grow and are dominant relative to floating-point costs. The increase in savings of relative communication costs of the IMEX method over SE in the regime of large timesteps and large number of nodes is made possible by the ability to subcycle the HO component within an implicit treatment of the LO component.

5 Conclusions

In this manuscript, we have discussed how the traditional semi-implicit and split-explicit methods can be cast in the framework of HOLO methods. In particular, the HOLO formulation can be used as a physics-based preconditioner for a fully-implicit method, allowing timesteps larger than those restricted by the gravity wave in explicit methods, and allowing the method to scale as that of explicit methods, but with larger implicit timesteps and second-order accuracy. Additionally, the HOLO formulation, when cast as an IMEX algorithm, is algorithmically scalable and second-order accurate with serial performance comparable to the split-explicit method. Moreover, in parallel and with an effective communication staging strategy and subcycling of the HO problem, the IMEX approach results in more than two orders of magnitude decrease in relative communication cost versus the split-explicit approach, making IMEX ideally suited for exascale environments.

Acknowledgements

This work was sponsored by the Los Alamos National Laboratory Directed Research and Development Program and by the US Department of Energy Office of Science. This work was performed under US government contract DE-AC52-06NA25396 for Los Alamos National Laboratory, which is operated by Los Alamos National Security, LLC, for the US Department of Energy (LA-UR-15-20576).

References

[1] Computational Co-Design for Multi-Scale Application in the Natural Sciences. `http://codesign.lanl.gov/projects/cocomans/`.

[2] A. Arakawa and V.R. Lamb. Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model. In J. Chang, editor, *Methods in Computational Physics*, volume 17, pages 173–265. Academic Press, New York, 1977.

[3] U. M. Ascher, S. J. Ruuth, and B. T. Wetton. Implicit-explicit methods for time-dependent partial differential equations. *SIAM J. Numer. Anal.*, 32(3):797–823, 1995.

[4] G. Chen, L. Chacón, and D. C. Barnes. An efficient mixed-precision, hybrid cpu–gpu implementation of a nonlinearly implicit one-dimensional particle-in-cell algorithm. *J. Comput. Phys.*, 231(16):5374–5388, 2012.

[5] G. Chen, L. Chacón, and D.C. Barnes. An energy-and charge-conserving, implicit, electrostatic particle-in-cell algorithm. *J. Comput. Phys.*, 230(18):7018–7036, 2011.

[6] J.K. Dukowicz and R.D. Smith. Implicit free-surface method for the Bryan-Cox-Semtner ocean model. *J. Geophys. Res.*, 99(C4):7991–8014, 1994.

[7] D. R. Durran and P. N. Blossey. Implicit-explicit multistep methods for fast-wave-slow-wave problems. *Mon. Weather Rev.*, 140(4):1307–1325, 2012.

[8] K.J. Evans, D.A. Knoll, and M.A. Pernice. Development of a 2-D algorithm to simulate convection and phase transition efficiently. *J. Comput. Phys.*, 219:404–417, 2006.

[9] R. Hallberg. Stable split time stepping schemes for large-scale ocean modeling. *J. Comput. Phys.*, 135(1):54–65, 1997.

[10] M.A. Heroux et al. An overview of the Trilinos project. *ACM Trans. Math. Softw.*, 31(3):397–423, 2005.

[11] R.L. Higdon and R.A. de Szoeke. Barotropic-baroclinic time splitting for ocean circulation modeling. *J. Comput. Phys.*, 135(1):30–53, 1997.

[12] S.Y. Kadioglu and D.A. Knoll. A fully second order implicit/explicit time integration technique for hydrodynamics plus nonlinear heat conduction problems. *J. Comput. Phys.*, 229(9):3237–3249, 2010.

[13] S.Y. Kadioglu, D.A. Knoll, R.B. Lowrie, and R.M. Rauenzahn. A second order self-consistent IMEX method for radiation hydrodynamics. *J. Comput. Phys.*, 229(22):8313–8332, 2010.

[14] J. Kim and P. Moin. Application of a fractional-step method to incompressible navier-stokes equations. *J. Comput. Phys.*, 59(2):308–323, 1985.

[15] D.A. Knoll and D.E. Keyes. Jacobian-free Newton-Krylov methods: a survey of approaches and applications. *J. Comput. Phys.*, 193(2):357–397, 2004.

[16] D.A. Knoll, H. Park, and K. Smith. Application of the jacobian-free newton-krylov method to nonlinear acceleration of transport source iteration in slab geometry. *Nucl. Sci. Eng.*, 167(2):122, 2011.

[17] J-F. Lemieux, D.A. Knoll, M. Losch, and C. Girard. A second-order accurate in time implicit-explicit (IMEX) integration scheme for sea ice dynamics. *J. Comput. Phys.*, 2014.

[18] M.E. Maltrud and J.L. McClean. An eddy resolving global 1/10 ocean simulation. *Ocean Model.*, 8(1):31–54, 2005.

[19] C. Newman and D.A. Knoll. Physics–based preconditioners for ocean simulation. *SIAM J. Sci. Comput.*, 35(5):S445–S464, 2013.

[20] C. Newman, G. Womeldorff, D.A. Knoll, and L. Chacón. An implicit-explicit subcycling method for a free-surface ocean model. *In prep*, 2015.

[21] R.C. Pacanowski, K. Dixon, and A. Rosati. The GFDL Modular Ocean Model Users Guide. Technical Report 2, Geophysical Fluid Dynamics Laboratory, Princeton, USA, 1993.

[22] H. Park, D.A. Knoll, and C.K. Newman. Nonlinear acceleration of transport criticality problems. *Nucl. Sci. Eng.*, 172(1):52, 2012.

[23] H. Park, D.A. Knoll, R.M. Rauenzahn, C.K. Newman, J.D. Densmore, and A.B. Wollaber. An efficient and time accurate, moment-based scale-bridging algorithm for thermal radiative transfer problems. *SIAM J. Sci. Comput.*, 35(5):S18–S41, 2013.

[24] H. Park, D.A. Knoll, R.M. Rauenzahn, A.B. Wollaber, and J.D. Densmore. A consistent, moment-based, multiscale solution approach for thermal radiative transfer problems. *Transp. Theory Stat. Phys.*, 41(3-4):284–303, 2012.

[25] T. Ringler, M. Petersen, R.L. Higdon, D. Jacobsen, P.W. Jones, and M. Maltrud. A multi-resolution approach to global ocean modeling. *Ocean Model.*, 69:211–232, 2013.

[26] R. Smith and P. Gent. Reference maunal for the Parallel Ocean Program (POP). Technical Report Los Alamos Technical Report LA-UR-02-2484, Los Alamos National Laboratory, 2002.

[27] W.T. Taitano, D.A. Knoll, L. Chacón, and G. Chen. Development of a consistent and stable fully implicit moment method for Vlasov–Ampère particle in cell (PIC) system. *SIAM J. Sci. Comput.*, 35(5):S126–S149, 2013.

[28] U. Trottenberg, C.W. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, 2000.

[29] W. Weijer, M.E. Maltrud, M.W. Hecht, H.A. Dijkstra, and M.A. Kliphuis. Response of the Atlantic Ocean circulation to Greenland Ice Sheet melting in a strongly-eddying ocean model. *Geophysical Research Letters*, 39(9), 2012.

[30] W. Ying, D.J. Rose, and C.S. Henriquez. Efficient fully implicit time integration methods for modeling cardiac dynamics. *IEEE Trans. Biomed. Eng.*, 55(12):2701–2711, 2008.